The Twenty-Fifth International Symposium on Artificial Life and Robotics 2020 (AROB 25th 2020),
The Fifth International Symposium on BioComplexity 2020 (ISBC 5th 2020),
B-Con Plaza, Beppu, Japan, January 22-24, 2020

Real-time Spatial Recognition by Underwater Stereo Vision

Takuro Kawakami[1], Renya Takahashi[1], Hongzhi Tian[1], Yejun Kou[1],Mamoru Minami[1]

[1]Okayama University, Japan
(Tel: 81-86-251-8233, Fax: 81-86-251-8233)
[1]p4zi5w9s@s.okayama-u.ac.jp

Abstract: Visual servoing is one of a method to control robot movement by installing the visual information into the feedback loop. So, it is expected to enable robots to work in a constantly changing environment or unknown environment. Even now, several methods have been studied to realize this. However, to recognize target object from images, some degree of predefined knowledge is needed. When it comes to using epipolar geometry, it's difficult to determine the corresponding point on images. Therefore, our research group proposed a new system that performs visual servoing that can avoid corresponding points identification problem. In our laboratory, we have been studying robots to explore the ocean. We would like to track new target sea animals with visual servoing in the ocean. In this study, we examine whether this method can be used in the dark by conducting experiments considering the dark environment such as the seabed.

Keywords: Visual servoing, Projection Method, 3D pose estimation,

1 INTRODUCTION

Currently, the aging is progressing in Japan. In addition, with the declining birthrate, it is expected that the declining birthrate and aging will progress in the future. For this reason, a decrease in the working population has become a major problem. Robots are expected to play an active role in various fields to compensate for the declining labor force. In addition, what is required of robots has changed from the one in which the operator works through the robot as before, to the one in which the robot itself takes autonomous action in a changing environment. Visual servo is one of the methods for controlling robots.

This is a method of incorporating visual information obtained from a camera or the like into a feedback loop. Hand-eye camera robots are widely used for bin picking as industrial robots because of the convenience of freely selecting the camera viewpoint. However, since hand-eye camera robots often use monocular hand-eye, there is a problem that the position / orientation recognition accuracy in the camera gaze depth direction is not good. For this reason, a recognition method combining a camera and a laser range finder has also been studied. However, there is a premise that the object is the same object, and if this is not satisfied, it will be erroneously recognized and cause the robot to malfunction. Similar problems occur in information processing that measures 3D position and orientation from compound-eye camera image information. On the other hand, the research on robot control based on visual servoing has proposed prediction technique of the moving matter using the movement models and nonlinear observers. However, there is a problem that it takes time to make the recognition error close to zero. Several years ago, our research group has proposed a method to recognize and handle targets with specified feature by applying a model-based matching method and genetic algorithm to clothing classification work. At present, we are focusing on recognition methods for targets in unknown environments that apply this technology. Object recognition by image processing is a technique for recognizing an object from acquired image information, and can grasp the position and shape of the object. Our group is also working on the development of a vehicle that searches the seabed. We thought that there are some creatures and resources that have not yet been discovered in the sea, and the purpose is to find them during the underwater search. In other words, the future issue is whether visual servoing can be performed on targets in unknown environments such as underwater. First, we use the manipulator to conduct experiments as shown in Fig.1. In this study, we verify the real-time position recognition and tracking experiments for simulated marine organisms using this method. The simulated marine organisms were installed in a pool filled with water, and experiments were conducted in an environment with steps and different brightness. I describe the position error and fitness of the recognition value.

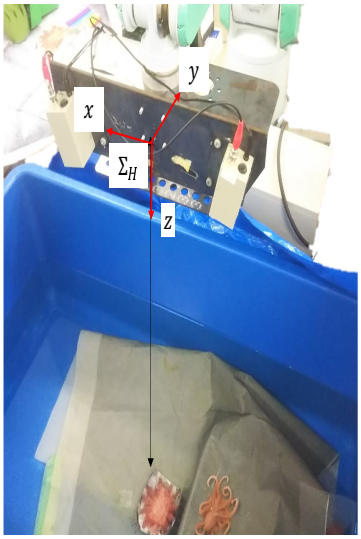

Fig. 1. Experiment environment

2 TARGET POSITION / POSTURE RECOGNITION METHOD

This chapter describes the projection-based matching method, the definition of the fitness function, and the three-dimensional position/posture measurement method using GA. This technique is based on the visual servo technology which is a previous research of this laboratory.

2.1 Projection-based Matching Method

The object recognition technique used in the robot system described in this paper requires still image recognition. Here, therefore, the outline of a recognition method for a single still image is described. Figure 2 shows a schematic diagram of the projection-based matching method.

The reference point is the hand coordinate system $\sum_H$. The reference coordinate system is $\sum_W$. The left camera coordinate system is $\sum_{CL}$, the image coordinate system is $\sum_{IL}$, the right camera coordinate system is $\sum_{CR}$, and the image coordinate system is $\sum_{IR}$. The target shown in the left camera image is extracted as a model, and the position / posture $\phi = ({}^{CL}z_{Mi}, {}^{Hx}\theta_M, {}^{Hy}\theta_M)$in 3D space is determined by the genetic algorithm (GA) gene. At this time, ${}^{CL}x_{MiC}$ and ${}^{CL}y_{MiC}$ are obtained from ${}^{CL}z_{MiC}$. The model is back-projected from the left camera image into the three-dimensional space. A two-dimensional planar model is obtained by reprojecting the image on the right camera. The fitness function is calculated and evaluated by matching the plane model with the right camera image. Then, when the position/posture ϕ (three variables) of the plane model coincides with the position/posture of the target, the value of the fitness function shows the maximum value.

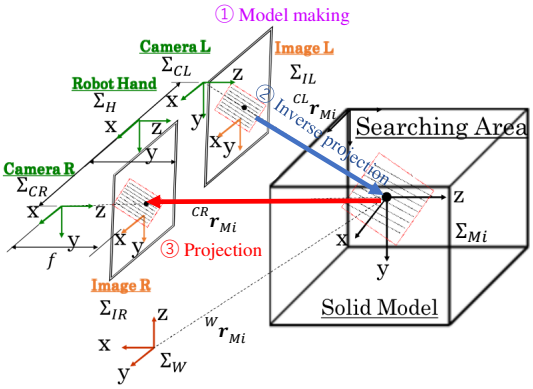

Fig. 2. Overview of Projection-based Method

2.2 Definition of fitness function

The image input from the left and right video cameras is composed of hue value represented by $0 \sim 359$ and brightness represented by $0 \sim 255$. The fitness function of hue value is shown in Eqs. (1)-(4). The fitness function of brightness value is shown in Eqs. (5)-(8). The search model consists of an inner region $p_{in}({}^I\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^j))$ and a outer region $p_{out}({}^I\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^j))$ to evaluate the target and the hue and brightness changes around the target.

The total number of search points in the inner area is N_{in}, and the total number of search points in the band area is N_{out}. The evaluation value based on the hue value in the right image region${}^{IR}r_i^j$is defined as $p_{R_H}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i))$. $D_H({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^j))$ is the difference in hue value between the model and the camera image. When the difference in hue value is within 20, " + 2 " is added as an evaluation value in the inner region and " 0.1 " is added in the outer region. If it is higher than this, " -1 " is added as an evaluation value in the inner area and " -1 " is added in the outer area.

$$F_{R_H}(\phi_M^i) = \left\{ \sum_{\substack{{}^{IR}r_i^j \in \\ S_{R,in}(\phi_M^i)}} p_{R_H,in}({}^{IR}r_i^j(\phi_M^i)) + \sum_{\substack{{}^{IR}r_i^j \in \\ S_{R,out}(\phi_M^i)}} p_{R_H,out}({}^{IR}r_i^j(\phi_M^i)) \right\} / (2 \times N_{R,in} + 0.1 \times N_{R,out}) \tag{1}$$

$$p_{R_H,in}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) = \begin{cases} 2, & \text{if}(D_H({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) \leq 20) \\ -1, & \text{otherwise.} \end{cases} \tag{2}$$

$$p_{R_H,out}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) = \begin{cases} 0.1, & \text{if}(D_H({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i))| \leq 20) \\ -1, & \text{otherwise.} \end{cases} \tag{3}$$

$$D_H({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) = |H_{IR}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) - H_{MR,in}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i))| \tag{4}$$

2.3 Fitness function of brightness value

The evaluation value based on the brightness value in the right image area ${}^{IR}r_i^j$is defined as $p_{R_B}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i))$. $D_B({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i))$ is the difference in brightness between the model and the camera image. The method of adding evaluation values is the same as the fitness function based on hue values.

$$F_{R_B}(\phi_M^i) = \left\{ \sum_{\substack{{}^{IR}r_i^j \in \\ S_{in}(\phi_M^i)}} p_{R_B,in} r_i^j(\phi_M^i)) + \sum_{\substack{{}^{IR}r_i^j \in \\ S_{out}(\phi_M^i)}} p_{R_B,out}({}^{IR}r_i^j(\phi_M^i)) \right\} / (2 \times N_{in} + 0.1 \times N_{out}) \tag{5}$$

$$p_{R_B,in}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) = \begin{cases} 2, & \text{if}(D_B({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) \leq 20) \\ -1, & \text{otherwise.} \end{cases} \tag{6}$$

$$p_{R_H,out}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) = \begin{cases} 0.1, & \text{if}(D_B({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i))| \leq 20) \\ -1, & \text{otherwise.} \end{cases} \tag{7}$$

$$D_B({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) = |B_{IR}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i)) - B_{MR,in}({}^{IR}\boldsymbol{r}_i^j(\boldsymbol{\phi}_M^i))| \tag{8}$$

2.4 Fitness function of hue and brightness value

The fitness function of hue and brightness value is shown in Eqs. (9). $F_{R_H}(\phi_M^i)$ and $F_{R_B}(\phi_M^i)$ are the degree of matching between the hue value and the brightness value described earlier.

$$F_R(\boldsymbol{\phi}_M^i) = F_{R_H}(\boldsymbol{\phi}_M^i) + F_{R_B}(\boldsymbol{\phi}_M^i) - (F_{R_H}(\boldsymbol{\phi}_M^i) \times F_{R_B}(\boldsymbol{\phi}_M^i) \tag{9}$$

The right image and the reprojected plane model are evaluated using the fitness function $F_R(\boldsymbol{\phi}_M^i)$. This fitness function $F_R(\boldsymbol{\phi}_M^i)$ is a function which uniquely determines the position/posture information of a model inverse-projected in space, and if the inverse-projected model coincides with the target, the target and the search model should also coincide in the right image. However, when $F_{R_H}(\boldsymbol{\phi}_M^i) \leq 0$, $F_{R_H}(\boldsymbol{\phi}_M^i)$= 0. when $F_{R_B}(\boldsymbol{\phi}_M^i) \leq 0$, $F_{R_B}(\boldsymbol{\phi}_M^i)$= 0.

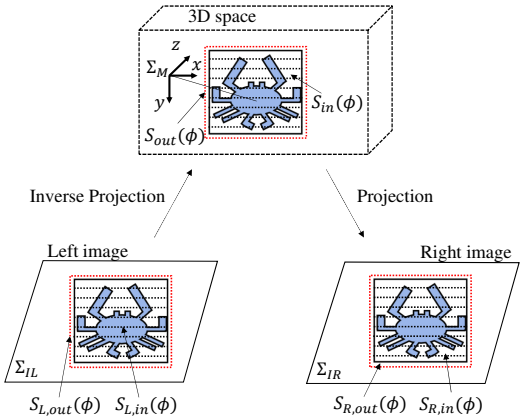

Fig. 3. Searching model

2.5 Optimal solution searching method using GA

There are various methods for searching and finding the maximum value of the fitness function, but the simplest and simplest method is the whole search method. In this method, the maximum value is found by calculating all function values, and the maximum value is always found, but it is inefficient. In other words, they have drawbacks such as spending a lot of computation time. The target visual servoing is a moving image recognition in real time, and it is important that the calculation processing is short for moving image recognition. In this study, by applying GA to optimum solution search, maximum value search processing was efficiently carried out in short time. Then, when GA converges to some extent, the position/posture of the target can be measured by considering the position/posture of the solid model determined by the gene of GA as the position/posture of the target. Therefore, real-time Multi-Step GA is used in this study. Figure 4 shows the search process using GA. This advances the GA for the input image only until the next image is input at the video rate (33 [ms]). It is a recognition method to output an individual giving the highest fitness at that time as a position and posture at that time.

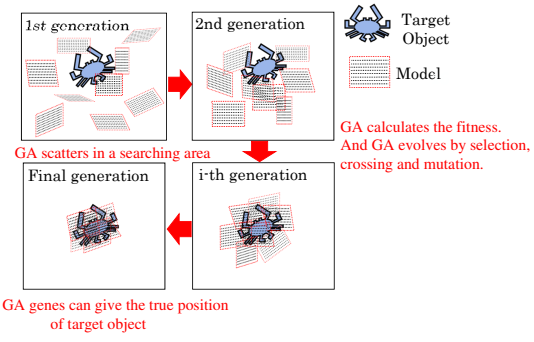

Fig. 4. Searching process using GA

3 EXPERIMENT

In this chapter, we conducted experiments to confirm that real-time position recognition is possible for underwater targets in dark environments with an illuminance of 5 [lux] or less. First, the experimental environment, experimental contents, and experimental results are described and discussed.

3.1 Fixed target experiment in the dark environment

3.1.1 Experiment environment

The actual sea area is in the difficult place to reach the dark light. Therefore, an experiment was conducted to see if the camera could be recognized even in dark conditions. Real-time recognition was performed in a dark environment with the target fixed in water. The distance from the hand coordinate system to the target was $(x, y, z) = (-40, 20, 840)[mm]$, and recognition was performed for 60 seconds. Figure 5 is an image captured by the left and right cameras. Position recognition is assumed to be possible if the average error of the X and Y coordinates is within 1/3 of the size of the target and the average error of the Z coordinates is within $50[mm]$. The size of Target in the X direction is $185[mm]$ and the size in the Y direction is $130[mm]$. The allowable error of Target is $(x, y, z) = (61, 43, 50)[mm]$ The reason why a large error is allowed only by the Z coordinate is that the recognition distance is more than $800[mm]$.

Fig. 5. Image of right camera

3.1.2 Experimental results and discussion

The experimental results are shown in Fig.6 ,7and 8 are the results of X, Y and Z, respectively. Figure 9 shows the result of the fitness value. The average of the measured values was $(x, y, z) = (-45, 20, 840)[mm]$. The average error was $(\Delta x, \Delta y, \Delta z) = (5, 2, 28)[mm]$. The average fitness value was 0.94.

From the experimental results, the average error of the X, Y, and Z coordinates are within the set target average error, so it is considered that the target can be recognized even in a dark environment.

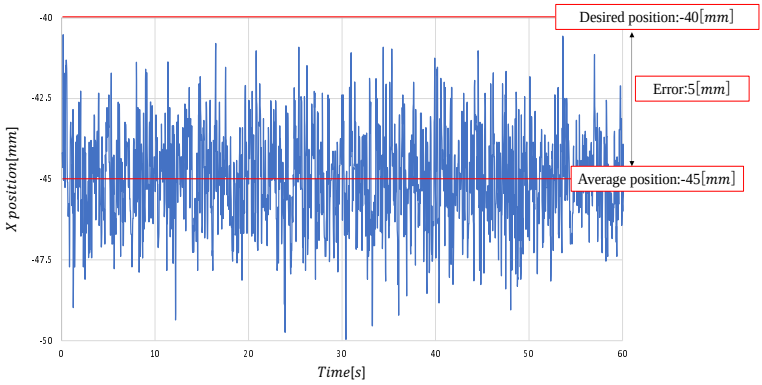

Fig. 6. Result of experiment in X

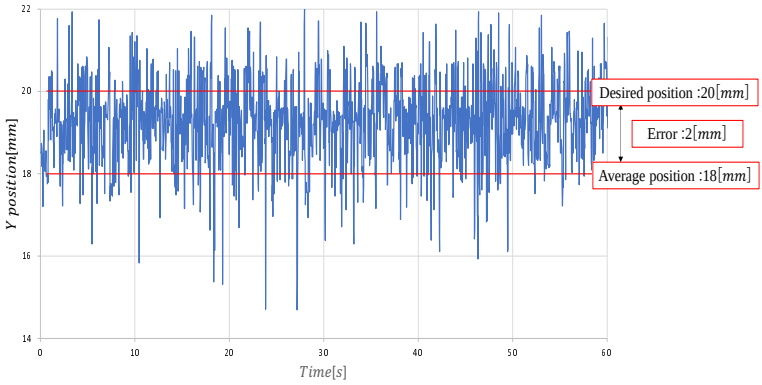

Fig. 7. Result of experiment in Y

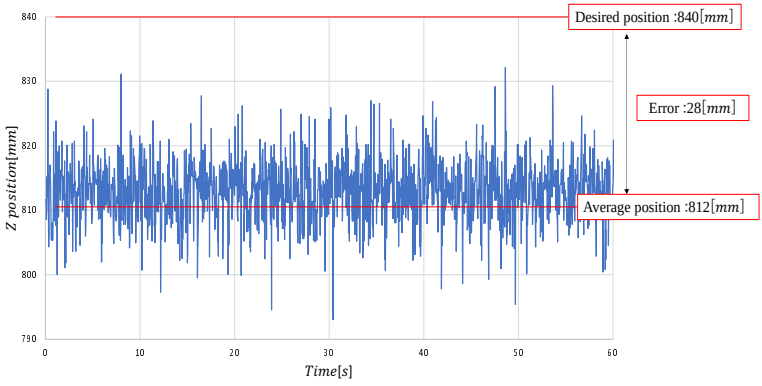

Fig. 8. Result of experiment in Z

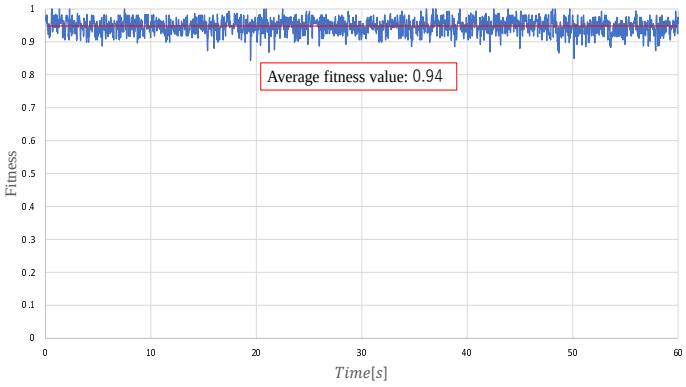

Fig. 9. Result of fitness

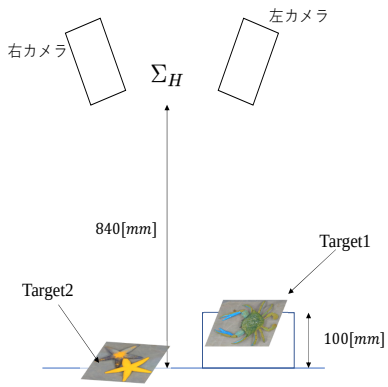

Fig. 10. schematic view

3.2 Height difference recognition experiment in dark environment

It was possible to recognize the fixed target on the plane.However, the natural environment is not a plane but an environment with a difference in elevation. Therefore, the next experiment was conducted in a dark environment with a height difference.

3.2.1 Experiment environment

Figure 10 outlines the experimental environment. The targets are Target 1 (crab) and Target 2 (starfish). Recognition from Target 1 to Target 2 was performed twice and position data was output. The coordinates of Target 1 are $(x, y, z) = (-170, 0, 740)[mm]$. The coordinates of Target 2 are $(x, y, z) = (170, -80, 840)[mm]$. As in the previous experiment, position recognition is assumed to be possible if the average error of the X and Y coordinates is within 1/3 of the size of the target and the average error of the Z coordinates is within $50[mm]$. The size of Target 1 in the X direction is $185[mm]$ and the size in the Y direction is $130[mm]$. The size of Target 2 in the X direction is $73[mm]$ and the size in the Y direction is $73[mm]$. That is, the allowable error of Target 1 is $(x, y, z) = (61, 43, 50)[mm]$. The allowable error of Target 2 is $(x, y, z) = (24, 24, 50)[mm]$.

3.2.2 Experimental results and discussion

The experimental results are shown in Fig.11 and Fig.12. The first measured value of Target 1 was $(x, y, z) = (-153, -13, 738)[mm]$, and the second time was $(x, y, z) = (-163, -23, 743)[mm]$. The average error from the measured values was $(\Delta x, \Delta y, \Delta z) = (17, 13, 2)[mm]$ for the first time. The second time was $(\Delta x, \Delta y, \Delta z) = (7, 23, 3)[mm]$. The first measured value of Target 2 was $(x, y, z) = (153, -79, 795)[mm]$ and the second was $(x, y, z) = (146, -82, 793)[mm]$. The average error from the measured values was $(\Delta x, \Delta y, \Delta z) = (17, 1, 46)[mm]$ for the first time. The second time was $(\Delta x, \Delta y, \Delta z) = (24, 2, 47)[mm]$.

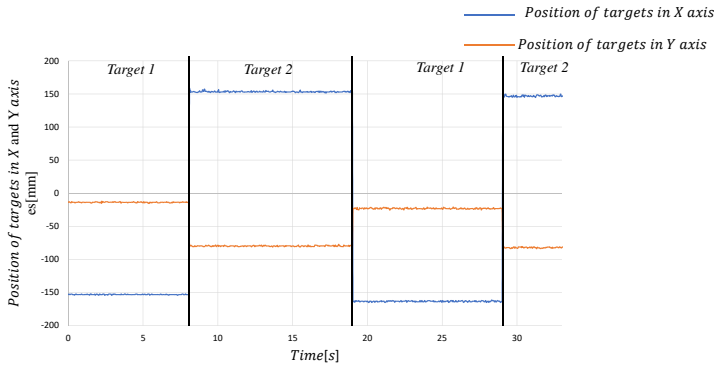

Fig. 11. The result of target1 and target2 position in X-Y axises

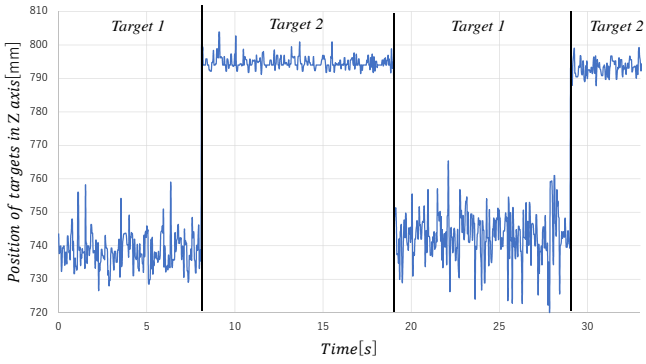

Fig. 12. The result of target1 and target2 position in Z axises

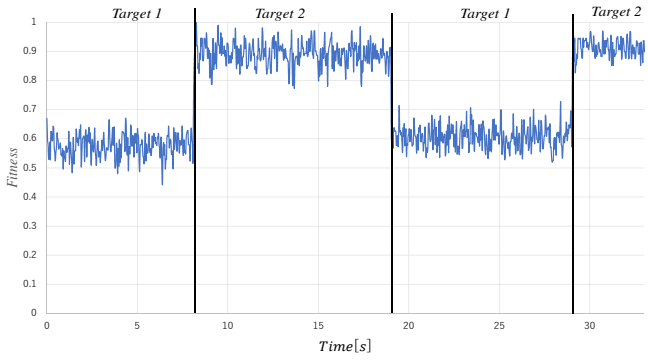

Fig. 13. The result of target1 and target2 fitness

From the experimental results, the average error of the X, Y, and Z coordinates was within the set target average error in each of Target 1 and Target 2. As can be seen from Fig.12, Target 1 has more vibration than Target 2. The fitness of Target 1 and 2 is shown in Fig.13. As can be seen from Fig.13, Target 2 has better the fitness than Target 1. Therefore, it is considered that the vibration at the Z-axis position is larger than Target 1. From these results, it is considered that the targets can be recognized even in the dark environment with height difference.

4 CONCLUSION

In this paper, it can be proven that real-time recognition under dark environment by this method is possible. Even if the target has almost no hue value, the target can be recognized with a small amount of light. This indicates that it can be adapted to unknown environments such as the seabed. In order to conduct experiments in actual sea areas using the underwater vehicle, we are considering the recognition of the environment considering turbidity as a future issue.

REFERENCES

[1] A.Lazaro, I.Serrano, J.P.Oria, “Ultrasonic circular inspection for object recognition with sensor-robot integration”, Sensors and Actuators A: Physical, Vol.77, Issue 1, pp.1-8, 1999.

[2] Yutaka Ishiyama, Yasushi Sumi,Fumiaki Tomita,“3-D Motion Lacking of 3-D Objects Using Stereo Vision”,Journal of the Robotics Society of Japan,Vol.18,No.2,pp213-220,2000

[3] Yoshihiro Nakabo, Idaku Ishii, Masatoshi Ishikawa,“3D Tracking using Two High Speed Vision Systems”,Intelligent Robots and Systems ,pp.360-365

[4] Sadaaki Yamane, Masao Izumi, Kunio Fukunaga, ‘‘A Method of Model-Based Pose Estimation ”, IEICE, Vol.J79-D-2, No.2,pp.165-173, Feb, 1996

[5] Fubito Toyama, Kenji Syoji,Juichi Miyamichi, ‘‘Pose Estimation from a Line Drawing Using Genetic Algorithm ”, IEICE, Vol.J81-D-2, No.7, pp.1584-1590, July, 1998

[6] S.Hutchinson, G.Hager, and P.Corke, “A Tutorial on Visual Servo Control”, IEEE Trans. on Robotics and Automation, vol. 12, no. 5, pp. 651-670, 1996.

[7] P.Y.Oh, and P.K.Allen, “Visual Servoing by Partitioning Degrees of Freedom”, IEEE Trans. on Robotics and Automation, vol. 17, no. 1, pp. 1-17, 2001.

[8] E.Malis, F.Chaumentte and S.Boudet, “2-1/2-D Visual Servoing”, IEEE Trans. on Robotics and Automation, vol. 15, no. 2, pp. 238-250, 1999.

[9] A. De Luca, G. Oriolo and P. R. Giordano, “On-line Estimation of Feature Depth for Image-Based Visual Servoing Schemes”, IEEE Int. Conf. on Robotics and Automation (ICRA2007).

[10] S.Caccavale, C.Natale, B.Siciliano and L.Villani, “Six-DOF Impedance Control Based on Angle/Axis Representations”, IEEE Trans. on Robotics and Automation, vol. 15, no. 2, APRIL 1999.

[11] B.Xian, M.S.de Queiroz, D.Dawson and I.Walker, “Task-Space Tracking Control of Robot Manipulators via Quaternion Feedback”, IEEE Trans. on Robotics and Automation, vol. 20, no. 1, FEBRUARY 2004.